

Weekly Report

July 1, 2018

1 Work

1. 降维论文本周又修改了一下程序，因为直接从100NN Graph构造层次结构的图误差会比较大，我们实验中发现1NN或者10NN构造的层次结构图比较有效。
2. Memory GAN的实验中，正在考虑Image和Text的特征如何结合的问题。我们对标的原始论文是直接把两个特征join起来，我们目前是用了一个Gate ($\text{Image} + \text{Gate} * \text{Text}$)。我看了一些VQA的论文，可以应用一些co-attention的方法。

1.1 工作进度

Table 1: 工作进度

项目	进度	截止时间
图布局方法扩展		7.30
降维	做数值实验，做性能比较	7.10
专利	完成撰写，等待律师回复	
AAAI投稿 (Memory GAN)	新增FID指标，但性能离原文章还有所差距	9.1

2 Paper Reading

2.1 Wasserstein GAN

原始的GAN的目标函数是JS divergence，它的梯度有时候会非常小，因此难以优化。WGAN提出Earth-Mover (EM) distance的目标函数，使得GAN的loss下降更快。

2.2 Improved Training of Wasserstein GANs

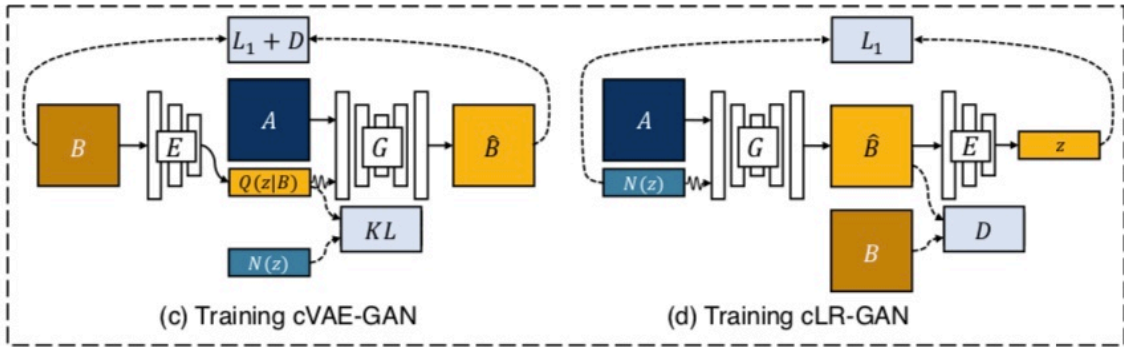
WGAN的loss对于神经网络有lipschitz约束，使得实际训练网络的时候要对参数做裁剪。WGAN-GP把这个约束放到目标函数中，使得训练更加方便。

$$L = \underbrace{\mathbb{E}_{\tilde{x} \sim \mathbb{P}_g} [D(\tilde{x})] - \mathbb{E}_{x \sim \mathbb{P}_r} [D(x)]}_{\text{Original critic loss}} + \underbrace{\lambda \mathbb{E}_{\hat{x} \sim \mathbb{P}_{\hat{x}}} [(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2]}_{\text{Our gradient penalty}}.$$

Figure 1: WGAN-GP

2.3 Toward Multimodal Image-to-Image Translation

文章将VAE的网络模块和GAN的模块结合起来，结合两个网络的优势训练出更加真实的图片。



2.4 Text2Shape: Generating Shapes from Natural Language by Learning Joint Embeddings

本文的原理是将文本和图片嵌入到同一个空间，这样可以根据文字很容易生成对应的图片。

2.5 Video Generation from Text

从文字生成图片，文章首先使用一个CVAE根据生成一个gist（初始内容图片），然后利用Text信息再次优化生成对应的视频。

